

Quantification of accuracy in field-based land cover maps: A new method to separate different components

Hannah E. E. S. Haga¹ | Anne B. Nilsen^{2,3} | Heidrun A. Ullerud²  | Anders Bryn^{2,3} 

¹SWECO, Hamar, Norway

²Geo-Ecology Research Group, Natural History Museum, University of Oslo, Oslo, Norway

³Division of Survey and Statistics, Norwegian Institute of Bioeconomy Research (NIBIO), Ås, Norway

Correspondence

Anders Bryn, Natural History Museum, University of Oslo, Oslo, Norway.
Email: anders.bryn@nhm.uio.no

Funding information

This research was funded by the Norwegian Biodiversity Information Centre, Trondheim, Norway.

Co-ordinating Editor: Duccio Rocchini

Abstract

Aim: Many thematic land cover maps, such as maps of vegetation types, are based on field inventories. Studies show inconsistencies among field workers in such maps, explained by inter-observer variation in classification and/or spatial delineation of polygons. In this study, we have tested a new method to assess the accuracy of these two components independently.

Location: Four study sites dominated by different ecosystems in southeast Norway.

Methods: We have used a vegetation-based land cover classification system adapted to a map scale of 1:5,000. First, a consensus map, a map that can be considered an approximation of a flawless map, was established. Secondly, the consensus map was adapted to test the accuracy of classification and polygon delineation independently. We used 10 field workers to generate a consensus map, and 14 new field workers (in pairs) to test the accuracy ($n = 7$).

Results: The results show that the accuracy of polygon delineation is lower than that of land cover classification. This is in contrast with previous studies, but previous research designs have not enabled a separation of the two accuracy components.

Conclusion: We recommend strengthening the training and harmonization of field workers in general, and increasing the emphasis on polygon delineation.

KEYWORDS

classification, ecosystem types, field survey, GIS, inconsistency, land cover mapping, map quality, Norway, vegetation mapping

1 | INTRODUCTION

The earth's surface is changing rapidly. There is high pressure on resources from increased land use, urbanization and climate change (Fuchs et al., 2015). Loss of pristine nature affects biodiversity, climate, soil stability, water circulation and groundwater reservoirs (Biondi et al., 2004). To protect nature or maintain a sustainable resource use, we need to know the distribution and condition of the present vegetation, as well as the impact of natural or human disturbances.

Land cover mapping is often the starting point for management planning or research purposes (Cherrill & McClean, 1999b; De Cáceres & Wiser, 2012). Land cover maps that include vegetation features are a good source for retrieving complex ecological information for a specific geographical area (Bryn et al., 2018). Such maps are based on knowledge and practice from two fields of applied research; botanical ecology and landscape geography (Küchler & Zonneveld, 1988).

Land cover maps depict the physical cover of the earth, and some classes are usually described by classification of vegetation

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Applied Vegetation Science* published by John Wiley & Sons Ltd on behalf of International Association for Vegetation Science

(Aune-Lundberg & Strand, 2017). Typically, vegetation is classified according to specific physiognomic features (Ihse, 2007) or characteristic groups of species that are found in locations with similar growing conditions (Box & Fujiwara, 2013). Many classification systems of land cover, outside strongly human-disturbed systems, capture more or less stable entities of either plant communities or ecosystems that re-appear in specific parts of the ecological space. These are usually characterized by species composition, physiognomy, indicator species or a combination of the three (Bryn et al., 2018). Other criteria that define land cover classes, besides vegetation, can be classes affected by human disturbances (for instance infrastructure, buildings, etc.) or natural disturbances (for instance landslides). Land cover classification systems are often hierarchical, where similar vegetation, ecosystems or other kinds of land cover are generalized into classes on different levels within a hierarchy (Cherrill & McClean, 1999a).

Land cover mapping of mutually exclusive and predefined types can be done in the field using a field computer and aerial photos, by interpretation of aerial photos, by using a variety of supervised (*sensu lato*) remote-sensing techniques or by for example distribution modeling (Fassnacht et al., 2016; Horvath et al., 2019). Alternatively, land cover mapping can be done with unsupervised remote-sensing techniques or based on fuzzy membership, the latter exemplified by Rocchini (2010). Field-based land cover maps are made by identifying and mapping areas of homogenous land cover (spatial delineation), and by assigning these polygons to predefined types (classification). Classification systems (types) and map generalizations (delineation) should be pre-adapted to a specific resolution through a defined scale intended for the map series (Hearn et al., 2011).

Land cover maps need high quality in order to be trusted by end-users (Cherrill, 2016). Considerable numbers of land cover maps exist, but there is often limited information on the reliability and quality of these data (Cherrill & McClean, 1995, 1999b; Hearn et al., 2011). All classification and mapping methods lead to an artificial generalization of nature (Green & Hartley, 2000). Continuity, gradual changes over space and diffuse borders with mixed species composition, may lead field workers to make arbitrary polygon lines (Hearn et al., 2011). Furthermore, a land cover type defined by vegetation is an abstract ideal; any land cover type delineated as a polygon will therefore be an imperfect representation of reality (Pancer-Koteja et al., 2009). These, as well as other factors, can give rise to map inconsistencies (Küchler & Zonneveld, 1988; Cherrill & McClean, 1995).

In studies evaluating quality of land cover maps, the term “inconsistencies” is commonly used when comparing observers and assessing inter-observer variation, i.e., when two or more observers obtain different results (Morrison, 2016; Ullerud et al., 2018). Inconsistency is thus defined as the difference between land cover maps made by different mappers, when all other factors are kept constant. In this study, we use the term accuracy to assess the deviation between a consensus map regarded as flawless (a “true” reference map) and land cover maps made by single mappers. According to a number of

previous studies, the main inconsistencies in field-based land cover data can be summarized in two categories: classification inconsistencies and spatial inconsistencies (Cherrill & McClean, 1999a; Ullerud et al., 2018). In classification inconsistencies, observers delineate roughly the same location, but assign different land cover types. In spatial inconsistencies, the observers assign the same land cover type, but delineate polygon borders differently (Cherrill, 2013) or include/exclude polygons (Möisja et al., 2018). Distinguishing between these two broad classes of inconsistencies in field-based land cover maps is challenging. There will always be inconsistencies in maps, but it is important to know the nature and scale of the sources, so that efforts can be made to improve the quality.

Since land cover maps are more or less affected by subjective decisions made during field work, a reference (“true”) land cover map is needed to evaluate accuracy. To measure consistency among mappers the same area can be mapped independently by different mappers and the degree of similarity between maps can be calculated (Cherrill & McClean, 1999b). A number of studies have assessed consistency in maps (Cherrill & McClean, 1995, 1999b; Hearn et al., 2011; Ullerud et al., 2018), but none of them have estimated accuracy and separated the effects of classification from spatial delineation of polygons as independent components. Subjectivity leading to unequal numbers of polygons among (commission/omission) different mappers (Möisja et al., 2018), prevents a straightforward interpretation of inconsistencies resulting from classification and delineation. New studies are therefore needed to separate the causes of map inconsistencies and to quantify accuracy. The main objective of this study is to quantify accuracy in field-based land cover mapping between observers and to develop a new method that enables a separation of the main causes of inaccuracy, while excluding effects of omission (fewer polygons than needed according to the guidelines) and commission (more polygons than needed according to the guidelines). The study is designed to answer the following questions concerning field-based land cover maps: (a) how accurate is the classification; (b) how accurate is the spatial delineation of polygons; (c) what characterizes land cover types that are more often inaccurately mapped; and (d) are some ecosystems more accurately mapped than others, and if so, why?

2 | METHODS

2.1 | Study area

The study area is located at Ringsakerfjellet in Hedmark county, southeast Norway (Figure 1). The area is within the northern boreal vegetation zone with low winter temperatures, warm summers (mean annual temperature between 0 and 2°C) and annual precipitation 1,000–1,500 mm (Moen, 1999). Ringsakerfjellet is a large mountain plateau ranging from 700 m to 1,000 m a.s.l. (Rekdal et al., 2003). The landscape is mostly below the climatic forest limit, which is lowered by centuries of extensive summer dairy farming. The bedrock consists of metamorphic sandstone and scattered intrusions of

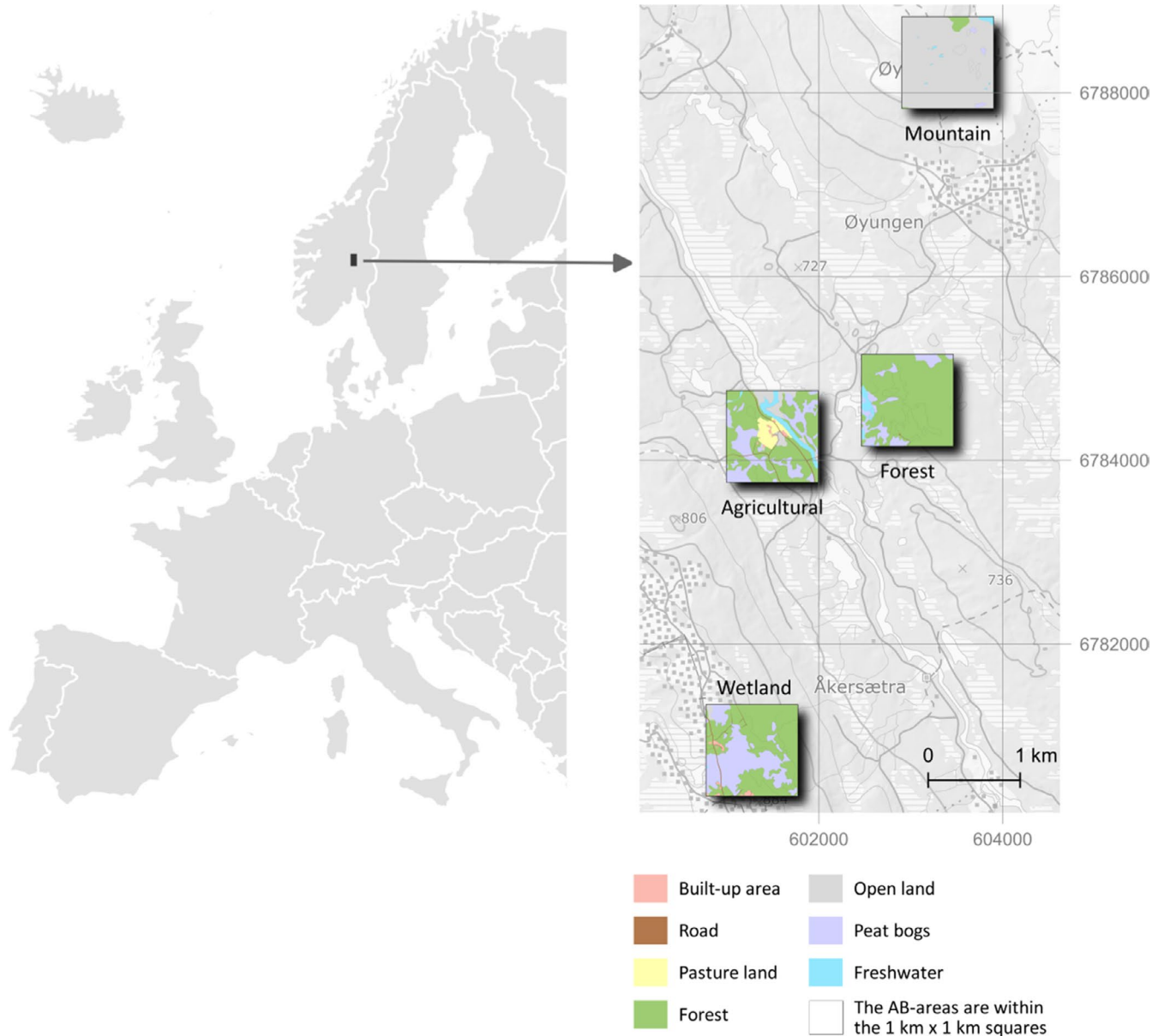


FIGURE 1 Location of study area and the four study sites in southeast Norway. Inset: Northern Europe with the study area marked (WGS 1984, UTM33)

lime-dominated bedrock. The soils are dominated by till, fluvial deposits and wetland (Rekdal et al., 2003). The area comprises forests characterized by birch woodland and stunted coniferous woodland, wetland, cultural landscape and scattered dwarf shrub-dominated mountains (Rekdal et al., 2003). Sheep and cattle graze in the outfields, and logging is common in the lower coniferous forests.

2.2 | Study design

Four rectangular study sites, each 50,000 m² and dominated by different ecosystems, were chosen for land cover mapping. The sites are named after the dominant ecosystem: mountain, agricultural, wetland and forest. The choice of sites was based on a vegetation

map from 2003 (Rekdal et al.), but the exact geographical location was determined in the field. The following criteria were considered: each site should preferably be dominated by one ecosystem, but include as much within-ecosystem variation as possible. The study consisted of two mapping parts, both following the official mapping guidelines for Norway by Bryn and Halvorsen (2015).

2.2.1 | Study design – part 1

In part one (Figure 2), the aim was to make consensus maps, approximations of flawless (“true”) reference maps for each study site. These were used to evaluate accuracy in classification and spatial delineation in part two of the study. The field work took

Parallel mapping by 10 field workers

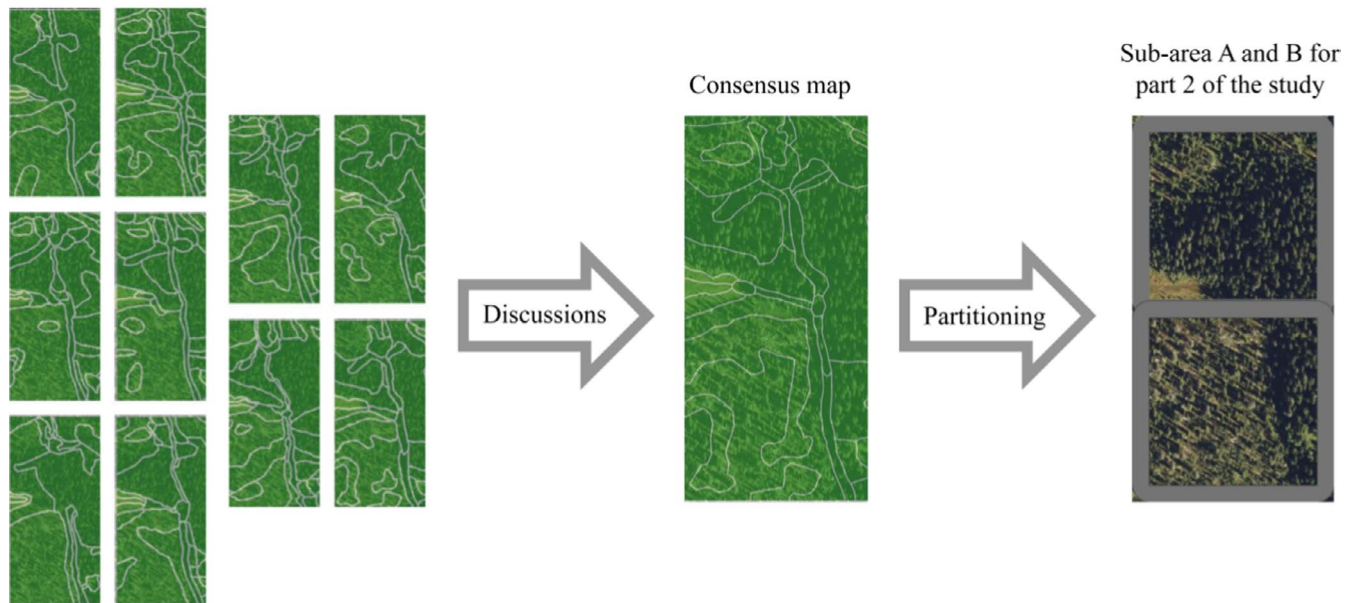


FIGURE 2 Study design of part 1. Parallel land cover maps made by ten field workers. The ten independent maps were then converted to one consensus map, and subsequently divided into two parts: A and B. The partitioning of the maps into A and B components forms the outset of the second part of the study design

place over five days in August 2017. Ten field workers were given an equal time-slot for practical mapping at each site, and everyone participated in the mapping of all sites. Afterwards, the most experienced field workers discussed the completed maps and prepared a first draft of the consensus maps. This draft was assessed and given comments and suggestions by all field workers. An improved set of drafts was sent out for approval by all participating field workers, and the consensus maps (one for each ecosystem) were completed after revisions.

The consensus maps were divided into two parts. Half of the maps included all the polygons from the consensus maps, but without information about the classification (sub-area A). The other half of the maps included classified points without polygon borders (sub-area B), one point for each original polygon from the consensus map.

2.2.2 | Study design – part 2

In part two, the aim was to investigate the accuracy of classification and spatial delineation, using another team of field workers. Mapping in part two was executed by 14 field workers divided into seven pairs. Half of the field workers involved were experienced mappers, whereas the other half consisted of master or PhD students. Pairs were put together so that the students were working together with experienced field workers. The field work took place over three days in September 2017. Sub-area A was mapped first by assigning land cover types (classification) to existing polygons. Thereafter sub-area B was mapped by delineating one polygon around each classified point with the aid of aerial photos, so that the result is wall-to-wall land cover maps.

2.2.3 | Training and calibration of field workers

To reduce inconsistency and obtain high-quality maps, field workers need to be harmonized by calibration and trained in advance (Ullerud et al., 2018; Eriksen et al., 2019). In this study, there were several training and calibration sessions before both parts of the study. The field workers were trained two weeks in advance, one week theory and one week field excursion. Information on the entire study area including bedrock, superficial deposits, ecological region, important species and current and historic land use of the area was given. Each field day started with a training and calibration session in the field, but outside the specific study sites. The training and calibration included recognizing indicators species and different land cover types, how to interpret topography and other landscape elements, aerial photo interpretation as well as other important factors that aid the distinction of land cover types and as background for robust spatial delineation of polygons.

2.2.4 | Mapping system

This study used a Norwegian classification system termed Nature in Norway (abbreviated NiN). The system has recently been translated and published internationally by Halvorsen et al. (2020). Only a short introduction of the system is provided here. NiN comprises three main dimensions; scale, land cover types and a variety of attributes. The system is, among other things, adapted to land cover mapping at a scale of 1:5,000. Division of types in NiN is based on how plants respond to environmental gradients, and the interval of ecological space they represent. The system is hierarchical and comprises three

levels (number of types in parentheses): major type group (7), major type (92) and basic type (741). The 448 basic types from wetland and terrestrial areas are aggregated into 281 land cover types, adapted to mapping at a scale of 1:5,000. Some of these land cover types (41) are defined by other criteria than species composition, for example land use or natural disturbances like rockslides. Land cover types are assigned to polygons by identifying the species composition. Each land cover type is described in the mapping guidelines for NiN (Bratli et al., 2017), including information about physiognomy, characteristic species, ecology, aerial photo characteristics, etc. These descriptions aid mappers in recognizing types during field work. The attribute system comprises complementary variables that can be used to add extra information that is not described by land cover types, for example, dominating tree species and percent tree cover. This study has not included any complementary variables from the attribute system.

2.3 | Field method

In both parts of the study, mapping was done in the field using portable field computers with QGIS version 2.18.14 (downloaded from QGIS Development Team, January 2018; <https://www.qgis.org/en/site/>) and aerial photos from 1973 (Series Ringsaker; 20 cm resolution; 16th June) and 2016 (Series Østlandet; 25 cm resolution; 3rd October). Field workers were equipped with field instructions (Bryn & Halvorsen, 2015), a graphical overview and descriptions of land cover types (Bratli et al., 2017; Bryn & Ullerud, 2017). Minimum polygon size was 250 m². In both parts of the study, field workers were not allowed to exchange information or compare their results while mapping.

2.4 | Data management and corrections

Data management and analysis were done in QGIS, Excel and R (downloaded from R Core Team, January 2018; <https://www.r-project.org/>). Maps from part one and two were corrected for technical errors (topology errors etc.).

2.4.1 | Accuracy in classification and spatial delineation

The accuracy was estimated in three ways and provided as percentages: (a) the pairwise comparison between each field worker and consensus; (b) the intersection between all field workers and consensus; and (c) the overall accuracy of classification and spatial delineation for each ecosystem (overall accuracy). The percentages provided for classification accuracy tested in sub-area A are calculated as the percentage correctly classified polygons. The percentages provided for spatial delineation accuracy tested in sub-area B are calculated as the percentage correctly delineated area (by intersect in GIS).

Ecological distance was used to quantify deviations in recorded land cover types relative to a reference, the consensus map (Figure 3). The ED between two types indicates to what degree they have a shared species pool (see Eriksen et al., 2019), i.e., how far apart the land cover types are within the larger ecological space. A higher ED indicates fewer species in common among the compared land cover types. When field workers have registered the same land cover type as consensus, the deviation is zero ED.

2.4.2 | Variation among ecosystems in mapping accuracy

Heat maps were constructed for each ecosystem to visually display the total mapping accuracy. Frequency of field workers that classified the same land cover type as consensus is represented by points with different colors. A point grid with 3-m spacing was used.

3 | RESULTS

A total of 56 maps were generated from part two of the study, 14 from each of the four sites, seven maps from each sub-area in all sites.

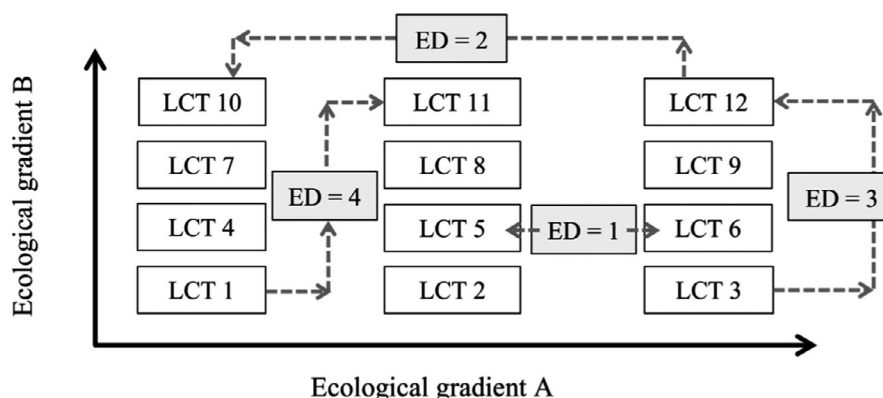


FIGURE 3 Examples of how ecological distance (ED) is calculated, based on the deviance between the consensus land cover type (LCT) and the registered land cover type

3.1 | Classification and spatial delineation accuracy

The pairwise comparison had higher accuracy than the overall comparison in both sub-areas (μ versus OA in lower section of Table 1). In the overall comparison, there was a higher accuracy in spatial delineation, with a mean overall classification accuracy of 30.5% and a mean overall spatial delineation accuracy of 33.1%. The results varied between different ecosystems in the overall comparison.

The mean pairwise classification accuracy is 72%, and the results from each ecosystem range from 55% in forest ecosystem to 97% in mountain ecosystem. Wetland ecosystem has the largest standard deviation. The mean classification accuracy for mountain ecosystem is significantly different from the mean classification accuracy of the three other ecosystems (Table 2).

The mean spatial delineation accuracy was 59%, ranging from 52% in agricultural ecosystem to 64% in wetland ecosystem. Agricultural ecosystem has the largest standard deviation. The wetland ecosystem has significantly different spatial delineation accuracy than the mountain and agricultural ecosystems.

3.2 | Ecological distance (ED)

The results in Table 3 reflect the same trends as the classification accuracy values in Table 1, where the forest ecosystem had the lowest accuracy (mean ED = 1) and the mountain ecosystem had the highest (mean ED = 0.4). The right-skewed frequency distribution of ED in all ecosystems showed that field workers chose land cover types that were ecologically related to consensus (Appendix S1). There is, however, a variation between ecosystems. In mountain ecosystem, as much as 86% of the observations had 0 ED, the rest of the

observations were spread from 1 ED to 6 ED. Wetland ecosystem displays the same pattern as mountain ecosystem with most of the observations (71%) having 0 ED from consensus. Forest and agricultural ecosystem show a more evenly distributed ED than the previous, and fewer observations have 0 ED from consensus, respectively 40% and 57%. Forest ecosystem had the largest number of registered land cover types and number of polygons (Table 3).

3.3 | Ecosystem complexity

Variation in mapping accuracy varies between ecosystems. Heat maps display the variation visually (Figures 4 and 5, remaining ecosystems given in Appendix S2). The least accurately classified land cover types (with 0 or 1 field worker pairs agreeing with consensus) are given in Appendix S3.

4 | DISCUSSION

4.1 | A new method to separate the main inaccuracies in mapping

Numerous studies have investigated the quality of field-based land cover maps and aimed to describe inconsistencies. In this study, however, we have developed a method to investigate the two main sources of inaccuracy separately; classification of land cover types vs spatial delineation of polygons. The results of implementing the AB partitioning show that in pairwise comparison between field workers and a consensus map, there was higher accuracy in classification than in spatial delineation. The mean classification accuracy

FW	Sub-area A					Sub-area B				
	M	A	W	F	μ	M	A	W	F	μ^1
1	100	54	49	36	60	47	58	65	62	58
2	100	74	53	67	74	54	29	48	57	47
3	93	77	80	74	81	58	86	74	70	72
4	94	69	90	42	74	42	32	51	65	48
5	97	39	38	48	55	59	53	72	68	63
6	96	78	100	81	87	53	53	73	59	60
7	96	72	78	36	71	64	53	63	57	59
μ	97	66	70	55	72	54	52	64	63	59
σ	2	13	22	17	11	7	17	10	5	8
CI	2	12	20	16	10	7	16	9	5	7
OA	88	1	33	0	31	27	35	40	30	33

TABLE 1 Classification accuracy from sub-area A and spatial delineation accuracy from sub-area B

Note: All pairs of field workers (FW) are compared with consensus. The following statistics is provided: mean accuracy (μ), standard deviation (σ) and confidence interval (CI; $\alpha = 0.05$). Results from each ecosystem; mountain (M), agricultural (A), wetland (W) and forest (F). The overall accuracy (OA) provides the result of all pairs of field workers compared with the consensus. All accuracy numbers are given in percentages.

TABLE 2 Paired significance test (Student's *t* test) of classification accuracy and spatial delineation accuracy with $\alpha = 0.05$ and Bonferroni-adjusted $\alpha = 0.0083$ for all combinations of ecosystems; mountain (M), agricultural (A), wetland (W) and forest (F)

	Sub-area A				Sub-area B			
	Classification accuracy				Spatial delineation accuracy			
	M	A	W	F	M	A	W	F
M								
A	0.002				0.775			
W	0.030	0.573			0.040	0.034		
F	0.001	0.115	0.151		0.060	0.142	0.720	

TABLE 3 Polygon characteristics from all ecosystems in sub-area A: Mountain (M), agricultural (A), wetland (W) and forest (F). Number of polygons (#), mean area of polygons ($\mu \text{ m}^2$) and mean ecological distance ($\mu \text{ ED}$)

Polygon	M	A	W	F
#	20	20	11	23
$\mu \text{ m}^2$	1,125	1,125	2,046	978
$\mu \text{ ED}$	0.4	0.9	0.5	1.0

was 72%, whereas the mean spatial delineation accuracy was 59%. This is in direct contrast to a number of studies that have concluded that classification is the main source of inconsistencies among maps made by different field workers (Cherrill & McClean, 1995, 1999a; Hearn et al., 2011; Ullerud et al., 2018). Cherrill and McClean (1995, 1999a) and Hearn et al. (2011) improved consistency by an average of only 4–5% when removing a buffer (10-m buffer in the study by Hearn et al., 2011) around the polygon delineations, thus concluding that classification is the main source of inconsistency. Their findings were based on buffering methods (Burrough et al., 2015), which we do not consider to be an independent evaluation method of classification vs spatial delineation inconsistencies. In our opinion, it is a measure of delineation precision (removal of delineation imprecision by buffering), rather than a full analysis of the complexity in spatial delineation of polygons in land cover maps. A full analysis is especially challenging in maps with low consistency, since this makes it even more difficult to separate classification and spatial delineation inconsistencies (Alexander & Millington, 2000). Since the method used in this study excludes the effects of omission and commission, we believe that our study is more purposeful when it comes to disentangling the effects of classification from spatial delineation. The results from our study show that the mean spatial delineation accuracy is much lower than in previous studies. The presented results indicate that the pairwise inconsistencies emerging from spatial delineation are larger than the inconsistencies emerging from classification alone. Consequently, field-based mapping programs should put more efforts into training and harmonizing spatial delineation of polygons.

The level of overall inconsistencies in this and comparable studies are approximately equal. Cherrill and McClean (1995, 1999a) and Hearn et al. (2011) found an overall consistency among field workers ranging from 25.6% to 34.2%, whereas the mean overall accuracy in this study is 30.5% for classification and 33.1% for spatial

delineation. Although not directly comparable, both results indicate that field-based land cover maps of types defined by vegetation (and land use) should be used with caution, particularly when implemented in monitoring programs or analyses of landscape changes (Bryn & Hemsing, 2012).

4.2 | Robustness with multiple field workers

In this study, 10 field workers' interpretation of the area is incorporated in the consensus map. This is not a perfect solution, but gives a more robust reference map than using only one field worker's map for comparison. Several vegetation studies recommend the use of multiple field workers, because working in teams has the effect of avoiding extreme estimates and detecting more species which are important to identify the land cover types (Symstad, Wienk & Thorstenson, 2008; Archaux, 2009; Gorrod & Keith, 2009; Vittoz et al., 2010). Ideally, one would use a map that depicts land cover in an unbiased and flawless way with absolute certainty. However, such field-based land cover maps do not exist, being generalizations of the real world (Zonneveld, 1989; Pancer-Koteja et al., 2009), and influenced by subjective judgments of the field workers (Hearn et al., 2011).

4.3 | Classification accuracy

There are many possible reasons for inaccuracies in classification. All classification methods result in maps with a degree of inaccuracy due to artificial simplification and generalization of natural features (Hearn et al., 2011). Multiscale phenomena, such as nature, vary in time and space. Selection of the most important aspects, when adapting characteristics of nature to a pre-defined scale, involves loss of information (Burrough & Frank, 1995). Classification accuracy also depends on the system involved. Ullerud et al. (2018) found that more complex classifications systems led to lower classification consistency among field workers, compared with less complex systems. Thus, classification inaccuracies are most likely scale- and system-dependent, and many studies are therefore not directly comparable. Our study documented a mean classification accuracy of 72% in the pairwise comparison between field workers and consensus map. Eriksen et al. (2019) found comparable levels using the same scale

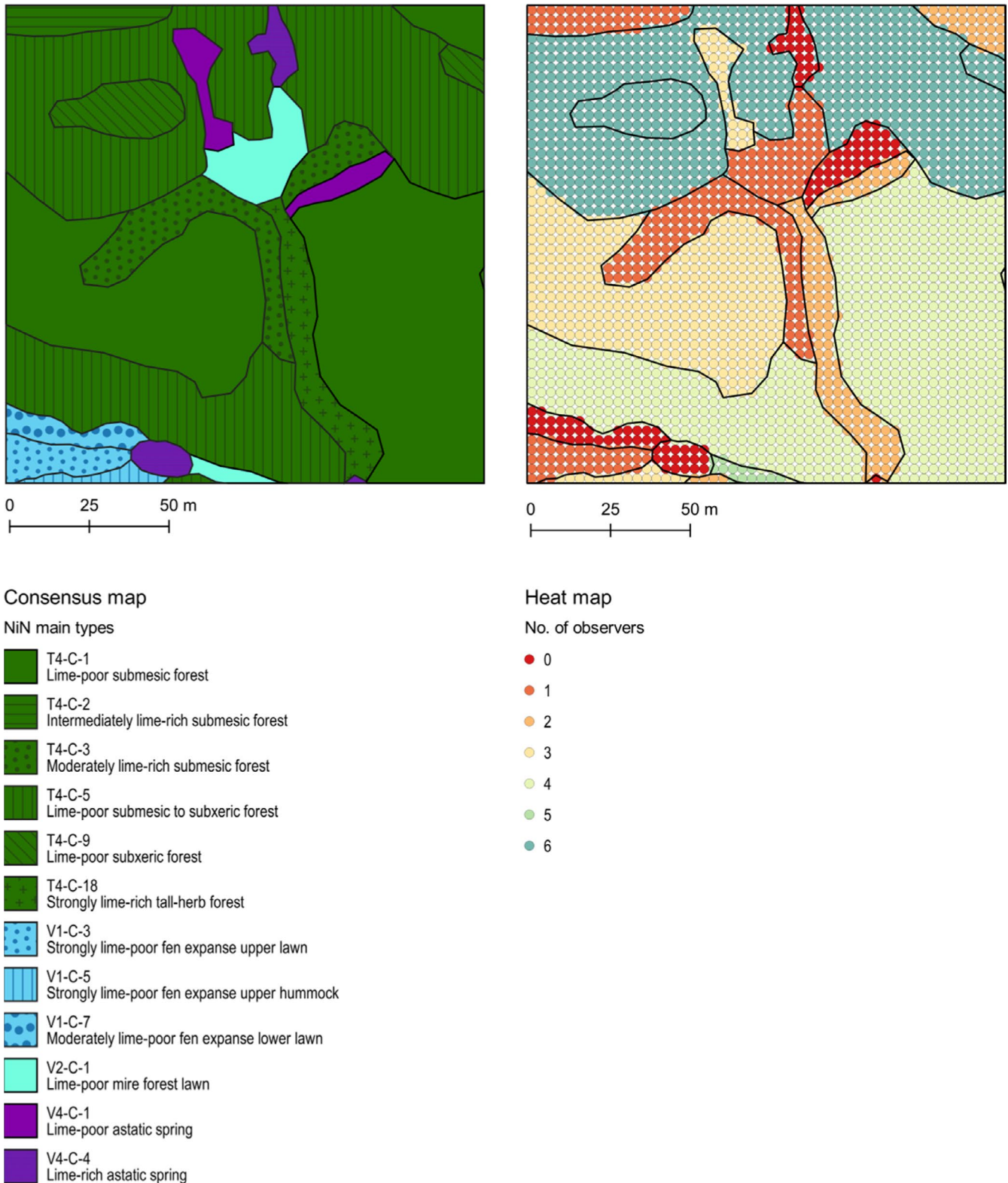


FIGURE 4 Comparison of consensus map (left) and heat map (right) from the forest ecosystem in sub-area A, visually displaying the classification accuracy. The heat map shows the number of field workers with similar classification as the consensus

and classification system, with a mean classification accuracy of 65%. Although Eriksen et al. (2019) used a point method (thereby avoiding delineation inaccuracy), it is reasonable to assume that

the mean level of classification accuracy varies around 55–75% for the tested land cover classification system (Ullerud et al., 2018; Eriksen et al., 2019; this study). The results are probably

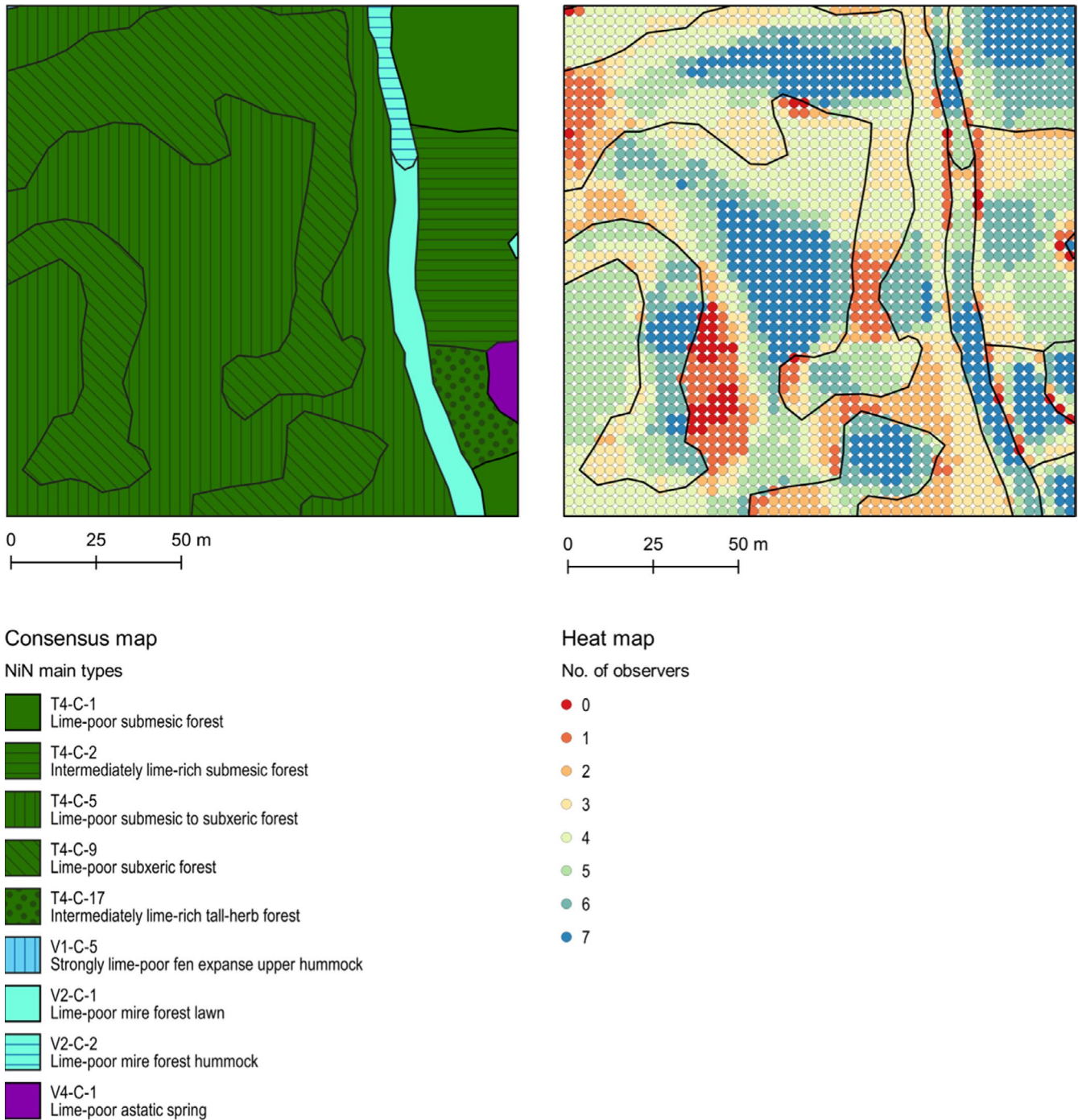


FIGURE 5 Comparison of consensus map (left) and heat map (right) from the forest ecosystem in sub-area B, visually displaying the spatial delineation accuracy. The heat map shows the number of field workers with similar spatial delineation as the consensus. Outline from consensus

context-dependent, and there are clear differences in classification accuracy between different ecosystems. Studies of field-based land cover mapping accuracy should therefore include multiple ecosystems.

Sufficient species knowledge is crucial in order to be able to recognize important indicator species needed to distinguish between land cover types. Varying ability to detect and identify species is a well-known cause of inconsistencies between field workers (Kirby,

2003; Bacaro et al., 2009; Hearn et al., 2011; Eriksen et al., 2019). Land cover types characterized by abundance of species that indicate a specific part of a gradient can also be challenging (Symstad et al., 2008). Regional and local variation of abundance can vary, and relative abundance of species can be troublesome to estimate correctly in field (Cherrill & McClean, 1999b). Gallegos Torrell and Glimskär (2009) recommend calibration with feedback to improve the accuracy of visual abundance estimates. Morrison (2016) found

that species can be overlooked and/or misidentified, where overlooking is a more prominent problem. Although Morrison's study is testing vegetation plots, similar challenges are likely to occur in the mapping of land cover types separated by differences in vegetation as well.

4.4 | Spatial delineation accuracy

Spatial delineation is well known to result in map inconsistencies among field workers (Cherrill & McClean, 1995, 1999a; Hearn et al., 2011; Ullerud et al., 2018), but has to our knowledge never been tested or reported as an independent component of field-based land cover maps in vector format. In our study, the mean spatial delineation accuracy is 59% with little variation between ecosystems (52–64%). The lowest accuracy is reported from the agricultural ecosystem, whereas the highest, in wetland ecosystem. Contrary to the classification accuracy, there is less variation in delineation accuracy between the ecosystems. Inaccuracy is overall high, although somewhat lower in wetlands. In sub-area B, field workers were given the specified land cover types at points. As expected, the field workers delineate consistently in the proximity of these points, but gradually less consistently with increasing distance from the points. The reported 58% mean spatial delineation accuracy is therefore probably a conservative estimate. If the points had been spatially randomized for each pair of field workers, the result would most likely end up with even lower spatial delineation accuracy.

Field-based land cover mapping is time-consuming and expensive. To map efficiently, the field workers use aerial photos for spatial delineation (Cherrill & McClean, 1999b; Ihse, 2007; Ullerud et al., 2020). Interpreting aerial photos requires experience and knowledge and relies on highly trained field workers (Morgan & Gergel, 2010). Fuzzy boundaries and more or less continuous vegetation (Couclelis, 1992), makes it difficult to delineate polygons. Even when borders between types are sharp and easy to interpret from aerial photos, the level of small-scale variation may be too complex for the intended map scale (Aune-Lundberg & Strand, 2017). Gradients in species cover, types defined by bottom and field layer species, moisture, soil nutrients, management level and succession state are considered the most difficult tasks to interpret from aerial photos, while separating open land from tree-covered areas is considered easier (Ihse, 2007). Our study, however, documents that types separated by a low estimate of species and tree cover boundary (e.g., above or below 10% tree crown cover) are difficult to spatially delineate. This is prominent along the boreal-alpine ecotone, and especially in areas influenced by land use that sustains a diffuse treeline (Harsch & Bader, 2011). Estimation of coverage is known to be difficult (Kennedy & Addison, 1987; Tonteri, 1990; van Hees and Mead, 2000) and shadows from trees can complicate the interpretation of aerial photos further (Ihse, 2007).

Although guided by aerial photo interpretation, inaccuracy in spatial delineation can also to some extent depend on the field workers' ability to distinguish adjacent types (Aspinall & Pearson, 1995). In this study, however, the land cover types were provided and therefore available for calibration before the spatial delineation. Our interpretation is therefore that this effect is negligible in this part of the study. Nevertheless, the lowest accuracy in spatial delineation is apparent between ecologically related land cover types and between strongly modified types that resemble semi-natural ecosystems.

4.5 | Land cover types with low accuracy

Accuracy varied with the land cover types that were mapped. This is also documented in other studies (Ullerud et al., 2018; Eriksen et al., 2019). ED, however, was typically low for land cover types with low classification accuracy. Land cover types that were most often confused were therefore ecologically closely related and always within the same major type. Others had high ED and were wrongly classified even at a higher hierarchical level (according to the consensus). Different major types can in some cases be very similar, with similar species composition, and mostly only separated for example by a scattered tree cover (above or below 10% crown cover) (Bratli et al., 2017), or differences in succession state and without distinct plant composition (Aune-Lundberg and Strand, 2017; Bratli et al., 2017). Land cover types separated by these attributes were frequently confused. Regrowth, late succession state and tree crown cover close to 10% can be the cause of this. Estimation of tree cover is challenging (Gallegos Torell & Glimskär, 2009) and the estimation is more difficult when tree cover is low (Morrison, 2016). In the implemented land cover system, species typical of semi-natural land cover types can gradually be replaced by species characteristic for forests, making such types challenging to classify correctly (Eriksen et al., 2019). This is comparable to other studies (Cherrill & McClean, 1999a; Hearn et al., 2011).

Land cover types within major types that were most often confused were often typically representing sections along gradients in lime richness, drought risk or rarity (also found by Eriksen et al., 2019). Field workers frequently chose land cover types with a lower lime richness than consensus. Although Eriksen et al. (2019) and Ullerud et al. (2018) reported opposite results, the cause might be the same. Classification inaccuracy among these types can indicate a lack of botanical skills needed to detect and recognize indicator species of lime richness and drought risk. Although not directly tested, we believe that the same challenges apply for low accuracy of rare land cover types as well.

Semi-natural land cover types were often confused with strongly modified types that resemble semi-natural ecosystems. Low accuracy in semi-natural land cover types seems to be common (Cherrill & McClean, 1999b; Stevens et al., 2004; Ullerud et al., 2018; Eriksen et al., 2019). Many land cover types in the tested system are defined

by land use or other strongly modified changes in addition to or instead of indicator species (Bratli et al., 2017). For mapping of land cover types defined by land use, extensive local knowledge or substantial experience is probably needed to make informed and correct classifications.

4.6 | Ecosystem complexity

The present results indicate that some ecosystems are more difficult to map consistently than others. Ecosystems with high numbers of land cover types had lower accuracy (forest and agricultural ecosystem) than ecosystems with fewer types. A higher number of available land cover types, with almost similar species composition, are therefore associated with lower accuracy (Cherrill & McClean, 1995, 1999a; Hearn et al., 2011; Ullerud et al., 2018). This implies that there is a trade-off between system complexity and map accuracy.

Forest ecosystem had the lowest accuracy with the largest deviation in ED from consensus (mean ED = 1). This is in accordance with other studies (Möisja et al., 2018; Ullerud et al., 2018). The used forest site had pronounced variation in topography, which may impact the results because of varying drought risk and lime richness (Ihse, 2007) leading to many possible land cover types to choose from, thus contributing to low accuracy (see sub-section 4.5 | *Land cover types with low accuracy*). Ullerud et al. (2018) also found low consistency in forest ecosystem when using the same mapping classification system (NiN). However, in the same study Ullerud et al. (2018) found the lowest consistency in wetland ecosystem when using another and coarser mapping system (Bryn et al., 2018). This is in contrast to the results from our study where wetland ecosystem had the highest classification accuracy. These results support Ullerud et al.'s (2018) hypothesis, that the classification system used for mapping may be more important for the resulting map consistency (and now accuracy), than the impact of different ecosystems.

4.7 | Uncertainties in this study

Field work is expensive and this study has a small sample size ($n = 7$). Although the mappers were working in pairs the statistical tests should be interpreted with caution. The number of polygons to be classified and points to be mapped, within each ecosystem, is also low. In addition, the results are from one area in Norway, tested with only one mapping system, and may therefore have limited transferability. In order to draw more certain conclusions, the study should be repeated elsewhere and with other mapping systems. The land cover classification system used in this study is also fairly new (2015), so at the time of this study there was not yet a pool of field workers with specific mapping experience. Therefore, in this test we had to blend in last-semester master students and PhD students (50%), together with more experienced mappers (50%). The inclusion of students might have lowered the resulting accuracy in the presented study,

although the use of pairs should prevent such effects. Anyway, both groups were trained for two weeks ahead of this study (see sub-section 2.2.3 | *Training and calibration of field workers*). Students are now extensively used for field survey campaigns in Norway, so the results of this study will be representative for the ongoing land cover mapping in Norway.

4.8 | Further studies

There are several measures that can be made to improve the quality of field-based land cover mapping. This study has, in our opinion, taken us a step closer to the understanding of the proximate causes of inaccuracy in mapping, but we have not investigated the ultimate causes, e.g., why some field workers are liable to delineate differently than others and what measures are most effective to improve classification accuracies. Improving the understanding of these underlying causes may help us to guide field workers better and could subsequently lead to reduced inconsistencies and higher accuracy.

In the presented study, effects of omission and commission are deliberately circumvented, but these effects are important to consider (Möisja et al., 2018; Ullerud et al., 2018). Omission and commission, however, can also be tested partly independently, so that the effects can be accounted for and compared with delineation and classification accuracy. We have started to plan a study targeting omission and commission, using a different design which allows field workers to define the number of potential polygons within an area where a consensus map is available. The new study, however, will be conducted the upcoming field season and reported thereafter.

5 | CONCLUSIONS

Pairwise comparisons show that the dominant source of inaccuracy is differences in spatial delineation. And, when deviating from consensus in assigning land cover types, ecologically closely related types are more frequently chosen. However, types that were defined by extensive land use (semi-natural types) or succession were more often misclassified as ecologically non-related types. There is variation among ecosystems when it comes to mapping accuracy, both in spatial delineation and classifications. Some ecosystems are more difficult to map than others.

We recommend that further work is carried out to determine ways of improving accuracy in field-based vector maps. Initial recommendations from this study are:

- Strengthen the training and harmonization of field workers in general, and increase the emphasis on polygon delineation
- In a land cover classification system with a high number of ecologically closely related types that are constantly mapped with low accuracy and consistency (in practice inseparable), these specific

classes should probably be merged when implemented for mapping purposes

- Develop and implement systematic mapping and monitoring of important quality parameters in land cover mapping programs
- Consider implementing the presented method, or comparable solutions, to pre-qualify field workers (certification) or to adapt the training according to the individual mappers' needs

ACKNOWLEDGEMENTS

Peter Horvath, Eva Lieungh Eriksen, Anders K. Wollan and Rune Halvorsen are acknowledged for providing scientific, technical or practical assistance.

AUTHOR CONTRIBUTIONS

HEESH, HAU and AB developed the idea and study design. They also conducted the research and wrote the manuscript. ABN developed the registrations schemes in QGIS, treasured the geodata and handled all GIS analyses. All authors discussed the results and commented on the manuscript.

DATA AVAILABILITY STATEMENT

The results are available in the manuscript and Appendix S2.

ORCID

Heidrun A. Ullerud  <https://orcid.org/0000-0003-0467-7089>

Anders Bryn  <https://orcid.org/0000-0003-4712-8266>

REFERENCES

- Alexander, R. & Millington, A.C. (2000) *Vegetation mapping: From patch to planet*. Chichester, UK: John Wiley & Sons LTD.
- Archaux, F. (2009) Could we obtain better estimates of plot species richness from multiple-observer plant censuses? *Journal of Vegetation Science*, 20(4), 603–611. <https://doi.org/10.1111/j.1654-1103.2009.01079.x>
- Aspinall, R. & Pearson, D.M. (1995) Describing and managing uncertainty of categorical maps in GIS. In: Fisher, P. (Ed.) *Innovations in GIS 2*. London: Taylor & Francis Ltd., pp. 71–83.
- Aune-Lundberg, L. & Strand, G.-H. (2017) Composition and spatial structure of dwarf shrub heath in Norway. *Norwegian Journal of Geography*, 71(1), 1–11. <https://doi.org/10.1080/00291951.2017.1291536>
- Bacaro, G., Baragatti, E. & Chiarucci, A. (2009) Using taxonomic data to assess and monitor biodiversity: Are the tribes still fighting? *Journal of Environmental Monitoring*, 11(4), 798–801. <https://doi.org/10.1039/b818171n>
- Biondi, E., Feoli, E. & Zuccarello, V. (2004) Modelling environmental responses of plant associations: A review of some critical concepts in vegetation study. *Critical Reviews in Plant Sciences*, 23(2), 149–156. <https://doi.org/10.1080/07352680490433277>
- Box, E.O. & Fujiwara, K. (2013) Vegetation types and their broad-scale distribution. In: van der Maarel, E. & Franklin, J. (Eds.) *Vegetation ecology*. Chichester: John Wiley & Sons Ltd., pp. 455–485.
- Bratli, H., Halvorsen, R., Bryn, A., Arnesen, G., Bendiksen, E., Jordal, J.B. et al. (Eds.) (2017) *Dokumentasjon av NiN versjon 2.1 tilrettelagt for praktisk naturkartlegging i målestokk 1:5000*. Trondheim: Norwegian Biodiversity Information Centre.
- Bryn, A. & Halvorsen, R. (2015) *Veileder for kartlegging av terrestrisk naturvariasjon etter NiN 2.0. Veileder versjon 2.0.0a*. Trondheim: Norwegian Biodiversity Information Centre.
- Bryn, A. & Hemsing, L.Ø. (2012) Impacts of land use on the vegetation in three rural landscapes of Norway. *International Journal of Biodiversity Science, Ecosystem Services & Management*, 8(4), 360–371. <https://doi.org/10.1080/21513732.2012.737373>
- Bryn, A., Strand, G.-H., Angeloff, M. & Rekdal, Y. (2018) Land cover in Norway based on an area frame survey of vegetation types. *Norwegian Journal of Geography*, 72(3), 131–145. <https://doi.org/10.1080/00291951.2018.1468356>
- Bryn, A. & Ullerud, H.A. (2017) *Veileder for arealdekkende terrestrisk naturvariasjon etter NiN i målestokk 1:5000 og 1:20000. Kort versjon (2.0.1.b)*. Trondheim: Norwegian Biodiversity Information Centre.
- Burrough, P.A. & Frank, A.U. (1995) Concepts and paradigms in spatial information: Are current Geographical Information Systems truly generic? *International Journal of Geographical Information Systems*, 9(2), 101–116. <https://doi.org/10.1080/02693799508902028>
- Burrough, P.A., McDonnell, R. & Lloyd, C.D. (2015) *Principles of geographical information systems*. Oxford: Oxford University Press.
- Cherrill, A. (2013) Repeatability of vegetation mapping using Phase 1 and NVC approaches: Implications for professional practice and surveyors' training requirements. *Practice*, 81, 41–45.
- Cherrill, A. (2016) Inter-observer variation in habitat survey data: investigating the consequences for professional practice. *Journal of Environmental Planning and Management*, 59(10), 1813–1832. <https://doi.org/10.1080/09640568.2015.1090961>
- Cherrill, A. & McClean, C. (1995) An investigation of uncertainty in field habitat mapping and the implications for detecting land cover change. *Landscape Ecology*, 10(1), 5–21. <https://doi.org/10.1007/bf00158550>
- Cherrill, A. & McClean, C. (1999a) Between-observer variation in the application of a standard method of habitat mapping by environmental consultants in the UK. *Journal of Applied Ecology*, 36(6), 989–1008. <https://doi.org/10.1046/j.1365-2664.1999.00458.x>
- Cherrill, A. & McClean, C. (1999b) The reliability of 'Phase 1' habitat mapping in the UK: the extent and types of observer bias. *Landscape and Urban Planning*, 45(2–3), 131–143. [https://doi.org/10.1016/s0169-2046\(99\)00027-4](https://doi.org/10.1016/s0169-2046(99)00027-4)
- Couclelis, H. (1992) People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS. In: Frank, A.U. (Ed.) *Theories and methods of spatio-temporal Reasoning in Geographic Space. Lecture notes in computer science*. Vol 639. Irene Campari and Ubaldo Formentini. Berlin, Heidelberg: Springer Verlag, pp. 65–77.
- De Cáceres, M. & Wiser, S.K. (2012) Towards consistency in vegetation classification. *Journal of Vegetation Science*, 23(2), 387–393. <https://doi.org/10.1111/j.1654-1103.2011.01354.x>
- Eriksen, E.L., Ullerud, H.A., Halvorsen, R., Aune, S., Bratli, H., Horvath, P. et al. (2019) Point of view: Error estimation in field assignment of land-cover types. *Phytocoenologia*, 49(2), 135–148. <https://doi.org/10.1127/phyto/2018/0293>
- Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T. et al. (2016) Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64–87. <https://doi.org/10.1016/j.rse.2016.08.013>
- Fuchs, R., Herold, M., Verburg, P.H., Clevers, J.G.P.W. & Eberle, J. (2015) Gross changes in reconstructions of historic land cover/use for Europe between 1900 and 2010. *Global Change Biology*, 21(1), 299–313. <https://doi.org/10.1111/gcb.12714>
- Gallegos Torell, Å. & Glimskär, A. (2009) Computer-aided calibration for visual estimation of vegetation cover. *Journal of Vegetation Science*, 20(6), 973–983. <https://doi.org/10.1111/j.1654-1103.2009.01111.x>
- Gorrod, E.J. & Keith, D.A. (2009) Observer variation in field assessments of vegetation condition: implications for biodiversity conservation. *Ecological Management & Restoration*, 10(1), 31–40. <https://doi.org/10.1111/j.1442-8903.2009.00437.x>
- Green, D.R. & Hartley, S. (2000) Integrating photointerpretation and GIS for vegetation mapping: some issues of error. In: Alexander, R. &

- Millington, A.C. (Eds.) *Vegetation mapping: From patch to planet*. West Sussex, UK: John Wiley & Sons Ltd., pp. 103–134.
- Halvorsen, R., Skarpaas, O., Bryn, A., Bratli, H., Erikstad, L., Simensen, T. et al. (2020) Towards a systematics of ecodiversity: the EcoSyst framework. *Global Ecology and Biogeography*, 29(11), 1887–1906. <https://doi.org/10.1111/geb.13164>
- Harsch, M.A. & Bader, M.Y. (2011) Treeline form – a potential key to understanding treeline dynamics. *Global Ecology and Biogeography*, 20, 582–596. <https://doi.org/10.1111/j.1466-8238.2010.00622.x>
- Hearn, S.M., Healey, J.R., McDonald, M.A., Turner, A.J., Wong, J. & Stewart, G.B. (2011) The repeatability of vegetation classification and mapping. *Journal of Environmental Management*, 92(4), 1174–1184. <https://doi.org/10.1016/j.jenvman.2010.11.021>
- van Hees, W.W.S. & Mead, B.R. (2000) Ocular estimates of understory vegetation structure in a closed *Picea glauca/Betula papyrifera* forest. *Journal of Vegetation Science*, 11(2), 195–200. <https://doi.org/10.2307/3236799>
- Horvath, P., Halvorsen, R., Stordal, F., Tallaksen, L.M., Tang, H. & Bryn, A. (2019) Distribution modelling of vegetation types based on an area frame survey. *Applied Vegetation Science*, 22(4), 547–560. <https://doi.org/10.1111/avsc.12451>
- Ihse, M. (2007) Colour infrared aerial photography as a tool for vegetation mapping and change detection in environmental studies of Nordic ecosystems: A review. *Norwegian Journal of Geography*, 61(4), 170–191. <https://doi.org/10.1080/00291950701709317>
- Kennedy, K.A. & Addison, P.A. (1987) Some considerations for the use of visual estimates of plant cover in biomonitoring. *Journal of Ecology*, 75(1), 151–157. <https://doi.org/10.2307/2260541>
- Kirby, K.J. (2003) NVC in semi-natural woodland. In: Goldberg, E. (Ed.) *National vegetation classification-ten years' experience using the woodland section*. JNCC Report 335. Peterborough: Joint Nature Conservation Committee, pp. 7–14.
- Küchler, A.W. & Zonneveld, I.S. (1988) *Vegetation mapping. Handbook of vegetation science*. Dordrecht: Kluwer Academic Publisher.
- Moen, A. (1999) *National atlas of Norway: Vegetation*. Hønefoss: The Norwegian Mapping Authority.
- Möisja, K., Uuemaa, E. & Oja, T. (2018) The implications of field worker characteristics and landscape heterogeneity for classification correctness and the completeness of topographical mapping. *ISPRS International Journal of Geo-Information*, 7(205), 1–16. <https://doi.org/10.3390/ijgi7060205>
- Morgan, J.L. & Gergel, S.E. (2010) Quantifying historic landscape heterogeneity from aerial photographs using object-based analysis. *Landscape Ecology*, 25(7), 985–998. <https://doi.org/10.1007/s10980-010-9474-1>
- Morrison, L.W. (2016) Observer error in vegetation surveys: a review. *Journal of Plant Ecology*, 9(4), 367–379. <https://doi.org/10.1093/jpe/rtv077>
- Pancer-Koteja, E., Szwagrzyk, J. & Guzik, M. (2009) Quantitative estimation of vegetation changes by comparing two vegetation maps. *Plant Ecology*, 205(1), 139–154. <https://doi.org/10.1007/s11258-009-9604-5>
- Rekdal, Y., Bryn, A. & Hofsten, J. (2003) *Vegetasjon og beite på Ringsakerfjellet. Rapport fra vegetasjonskartlegging. NIJOS Report Vol. 8 No 3*. Ås: Norwegian Institute of Land Inventory.
- Rocchini, D. (2010) While Boolean sets non-gently rip: A theoretical framework on fuzzy sets for mapping landscape patterns. *Ecological Complexity*, 7(1), 125–129. <https://doi.org/10.1016/j.ecocom.2009.08.002>
- Stevens, J.P., Blackstock, T.H., Howe, E.A. & Stevens, D.P. (2004) Repeatability of Phase 1 habitat survey. *Journal of Environmental Management*, 73(1), 53–59. <https://doi.org/10.1016/j.jenvman.2004.05.009>
- Symstad, A.J., Wienk, C.L. & Thorstenson, A.D. (2008) Precision, repeatability, and efficiency of two canopy-cover estimate methods in northern Great Plains vegetation. *Rangeland Ecology & Management*, 61(4), 419–429. <https://doi.org/10.2111/08-010.1>
- Tonteri, T. (1990) Inter-observer variation in forest vegetation cover assessments. *Silva Fennica*, 24(2), 189–196. <https://doi.org/10.14214/sf.a15575>
- Ullerud, H.A., Bryn, A., Halvorsen, R. & Hemsing, L.Ø. (2018) Consistency of land cover mapping; influence of fieldworkers, spatial scale and classification system. *Applied Vegetation Science*, 21(2), 278–288. <https://doi.org/10.1111/avsc.12368>
- Ullerud, H.A., Bryn, A. & Skånes, H. (2020) Bridging theory and implementation - Testing an abstract classification system for practical mapping by field survey and 3D aerial photographic interpretation. *Norwegian Journal of Geography*, 73(5), 301–317. <https://doi.org/10.1080/00291951.2020.1717595>
- Vittoz, P., Bayfield, N., Brooker, R., Elston, D.A., Duff, E.I., Theurillat, J.-P. et al. (2010) Reproducibility of species lists, visual cover estimates and frequency methods for recording high-mountain vegetation. *Journal of Vegetation Science*, 21(6), 1035–1047. <https://doi.org/10.1111/j.1654-1103.2010.01216.x>
- Zonneveld, I.S. (1989) The land unit – a fundamental concept in landscape ecology and its applications. *Landscape Ecology*, 3(2), 67–86. <https://doi.org/10.1007/bf00131171>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Appendix S1. Ecological distance (ED)

Appendix S2. Heat maps

Appendix S3. Least accurately classified land cover types

How to cite this article: Haga HEES, Nilsen AB, Ullerud HA, Bryn A. Quantification of accuracy in field-based land cover maps: A new method to separate different components. *Appl Veg Sci*. 2021;24:e12578. <https://doi.org/10.1111/avsc.12578>